WO 01/34796

PCT/IL00/00736

CHORDIN-LIKE HOMOLOGS

FIELD OF THE INVENTION

The present invention concerns novel nucleic acid sequences, vectors and host cells containing them, amino acid sequences encoded by said sequences, and antibodies reactive with said amino acid sequences, as well as pharmaceutical compositions comprising any of the above. The present invention further concerns methods for screening for candidate activator or deactivators utilizing said amino acid sequences.

BACKGROUND OF THE INVENTION

The TGFß superfamily is composed of a range of functional and structural factor subclasses with predominantly growth-inhibitory cellular actions and developmental regulatory effects on organogenesis, pattern formation, modulation of extracellular matrix and terminal differentiation. The subclasses include the TGFß, activins, glial-derived factors (GDFs), Mullerian inhibiting substances, glial-derived neurotrophic factor (GDNF), cartilage-derived morphogenetic proteins (CDMPs) and the rapidly expanding factor subclass of bone morphogenic proteins (BMPs). BMPs participate in a broad spectrum of cellular inducing events involving all three germ layers during metazoan development. There are now known to be 7 members of this family (BMPs 1-7); all except BMP1 are members of the TGF-α family. BMP1 has been classed as a novel regulatory protein. The term 'bone morphogenetic' may, however, prove to be a misnomer, since the messenger RNA for the BMPs are expressed in a wide variety of tissues, suggesting limited tissue specificity of function.

Chordin is an abundant glycoprotein with molecular mass of 120Kda. It contains internal cystein rich repeats called Von Willbrand domain and N-glycosylation sites.

Chordin is a key developmental protein that dorsalizes early vertebrate embryonic tissues by binding to ventralizing TGF-beta -like bone morphogenetic proteins (BMP) and sequestering them in latent complexes. Chordin binds to ventral BMP-2 and BMP-4 signal in the extracellular space, blocking the interaction of BMPs with their receptors. Chordin mimics the action of the Spemann organizer and can induce the formation of neural tissue from ectoderm and dorsalization of the ventral mesoderm to form muscle.

GLOSSARY

In the following description and claims use will be made, at times, with a variety of terms, and the meaning of such terms as they should be construed in accordance with the invention is as follows:

"Chordin like homolog (CLH) nucleic acid sequence" – the sequence shown in any one of SEQ ID NO: 1 to 11, sequences having at least 70% identity to said sequence and fragments of the above sequences being 20 b.p. long. Those sequences are sequences coding for a novel homolog of the known Chordin protein, as well as for variants of the novel homolog produced by alternative splicing.

The sequence shown in SEQ ID NO: 1 is a homolog to the known chordins within the VWFC domain, named after the von-Willebrand factor (VWF) type C repeat, which is found 2-4 times in these multi-domain proteins. The VWF domain has a length of about 70 amino acids covering 10 well conserved cysteines. The presence of this region in complex-forming proteins leads to the assumption that the VWFC domain might be involved in forming larger protein complexes. The homolog is a part of a longer sequence termed hereinafter *"full sequence"*. The full sequence has naturally occurring splice variants which are also termed CLH. The first variant (SEQ ID NO: 2) has 3 out of VWFC domains of the known

chordin. The protein coded therefrom contains a predicted signal peptide. The second variant (SEQ ID NO: 3) and third variant (SEQ ID NO: 4) contain 3 out of 4 VWFC domains of the known chordin, but is not predicted to contain the signal sequence. Sequences SEQID 5 to SEQ ID NO.10 are also splice variants of the full sequence. SEQ ID NO. 5 contains 3 out of 4 VWFC domains of known chordins (domain #2,3,4). The VWFC domain is named after the von Willebrand factor (VWF) type C repeat which is found twice in this multidomain protein. It has a length of about 70 amino acids covering 10 well conserved cysteines. The protein of chordin-like variant 1 at SEQ ID 5 (depicted in SEQ ID 16) contains predicted signal peptide. SEQ ID NO. 6 contains 3 out of the 4 VWFC domains of known chordin and the protein encoded thereby (SEQ ID 17), contains a predicted signal peptide. SEQ ID NO. 7 contains 2 out of 4 VWFC domains of known chordin. SEQ ID NO. 8 has out of the 4 VWF Factor Type C domains. SEQ ID NO. 9 has 2 VWF and SEQ ID NO. 10 has 2 VWF domains. SEQ ID No. 11 is a mouse ortolog of the CLH of the inventor.

However, the term CLH does not necessarily signify that CLH protein coded by the above sequences (including the variant sequences) has the same or even similar physiological effects as known Chordins, merely that it shows sequence homology with the known Chordin.

"Variant" – a sequence produced by alternative splicing of full sequence homolog. These sequences are not merely truncated forms of the full sequence, or modifications of the full sequence, but rather naturally occurring sequences resulting from various alternative splicings.

"Chordin like homology product (CLH product) – also referred at times as the "CLH protein" or ***"CLH polypeptide"*** – is an amino acid coded by any one of SEQ ID NOS: 1 to 11. The amino acid sequence may be a peptide, a protein, as well as peptides or proteins having *chemically modified* amino acids (see below) such as a glycopeptide or glycoprotein. An example of an CLH product is shown in any one of SEQ ID NOS: 12 to 22. The term also includes analogues of said sequences in which one or more amino acids has been added, deleted, *substituted*

(see below) or *chemically modified* (see below) as well as fragments of this sequence having at least 10 amino acids.

"Nucleic acid sequence" – a sequence composed of DNA nucleotides, RNA nucleotides or a combination of both types and may includes natural nucleotides, chemically modified nucleotides and synthetic nucleotides.

"Amino acid sequence" – a sequence composed of any one of the 20 naturally appearing amino acids, amino acids which have been *chemically modified* (see below), or composed of synthetic amino acids.

"Fragment of CLH product" - a polypeptide which has an amino acid sequence which is the same as part of but not all of the amino acid sequence of the CLH product.

"Fragments of CLH nucleic acid sequence" a continuous portion, preferably of about 20 nucleic acid sequences of the CLH nucleic acid sequence.

"Conservative substitution" - refers to the substitution of an amino acid in one class by an amino acid of the same class, where a class is defined by common physicochemical amino acid side chain properties and high substitution frequencies in homologous proteins found in nature, as determined, for example, by a standard Dayhoff frequency exchange matrix or BLOSUM matrix. [Six general classes of amino acid side chains have been categorized and include: Class I (Cys); Class II (Ser, Thr, Pro, Ala, Gly); Class III (Asn, Asp, Gln, Glu); Class IV (His, Arg, Lys); Class V (Ile, Leu, Val, Met); and Class VI (Phe, Tyr, Trp). For example, substitution of an Asp for another class III residue such as Asn, Gln, or Glu, is a conservative substitution.

"Non-conservative substitution" - refers to the substitution of an amino acid in one class with an amino acid from another class; for example, substitution of an Ala, a class II residue, with a class III residue such as Asp, Asn, Glu, or Gln.

"Chemically modified" - when referring to the product of the invention, means a product (protein) where at least one of its amino acid resides is modified either by natural processes, such as processing or other post-translational modifications, or by chemical modification techniques which are well known in the art. Among the numerous known modifications typical, but not exclusive examples include: acetylation, acylation, amidation, ADP-ribosylation, glycosylation, GPI anchor formation, covalent attachment of a lipid or lipid derivative, methylation, myristlyation, pegylation, prenylation, phosphorylation, ubiqutination, or any similar process.

"Biologically active" - refers to the CLH product which have, regulatory or biochemical functions on the same target sites which naturally occurring CLH influence, for example can bind to the same receptor as the chordin (or to another receptor).

"Immunologically active" defines the capability of a natural, recombinant or synthetic CLH product, or any fragment thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies. Thus, for example, a biologically active fragment of CLH product denotes a fragment which retains some or all of the immunological properties of the CLH product, e.g can bind specific anti-CLH product antibodies or which can elicit an immune response which will generate such antibodies or cause proliferation of specific immune cells which produce CLH.

"Optimal alignment" - is defined as an alignment giving the highest percent identity score. Such alignment can be performed using a variety of commercially

available sequence analysis programs, such as the local alignment program LALIGN using a ktup of 1, default parameters and the default PAM. A preferred alignment is the one performed using the CLUSTAL-W program from MacVector (TM), operated with an open gap penalty of 10.0, an extended gap penalty of 0.1, and a BLOSUM similarity matrix. If a gap needs to be inserted into a first sequence to optimally align it with a second sequence, the percent identity is calculated using only the residues that are paired with a corresponding amino acid residue (i.e., the calculation does not consider residues in the second sequences that are in the *"gap"* of the first sequence).

"Having at least X% identity" - with respect to two amino acid or nucleic acid sequence sequences, refers to the percentage of residues that are identical in the two sequences when the sequences are optimally aligned. Thus, 70% amino acid sequence identity means that 70% of the amino acids in two or more optimally aligned polypeptide sequences are identical.

"Isolated nucleic acid molecule having an CLH nucleic acid sequence" - is a nucleic acid molecule that includes the coding CLH nucleic acid sequence. Said isolated nucleic acid molecule may include the CLH nucleic acid sequence as an independent insert; may include the CLH nucleic acid sequence fused to an additional coding sequences, encoding together a fusion protein in which the CLH coding sequence is the dominant coding sequence (for example, the additional coding sequence may code for a signal peptide); the CLH nucleic acid sequence may be in combination with non-coding sequences, e.g., introns or control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host; or may be a vector in which the CLH protein coding sequence is a heterologous.

"Expression vector" - refers to vectors that have the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are known and/or commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

"Deletion" - is a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.

"Insertion" or "addition" - is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to the naturally occurring sequence.

"Substitution" - replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively. As regards amino acid sequences the substitution may be conservative or non- conservative.

"Antibody" – refers to IgG, IgM, IgD, IgA, and IgG antibody. The definition includes polyclonal antibodies or monoclonal antibodies. This term refers to whole antibodies or fragments of the antibodies comprising the antigen-binding domain of the anti-CLH product antibodies, e.g. antibodies without the Fc portion, single chain antibodies, fragments consisting of essentially only the variable, antigen-binding domain of the antibody, etc.

"Activator" - as used herein, refers to a molecule which mimics the effect of the natural CLH product or at times even increases or prolongs the duration of the biological activity of said product, as compared to that induced by the natural product. The mechanism may be by binding to the same receptor of target moieties to which native CLH binds thus mimicking the activity of CLH; by prolonging the lifetime of the CLH, (for example by decrease of the rate of its

degradation) by increasing the activity of the CLH on its target (modulation of expression and amount of BMPs), by increasing the affinity of CLH to moieties which it binds (such as its receptors) etc. Activators may be polypeptides, nucleic acids, carbohydrates, lipids, or derivatives thereof, or any other molecules which can bind to and activate the CLH product.

"Deactivator" or ***("Inhibitor")*** - refers to a molecule which modulates the activity of the CLH product in an opposite manner to that of the activator, by decreasing or shortening the duration of the biological activity of the CLH product. This may be done by blocking the binding of the CLH to its receptor (competitive or non-competitive inhibition), by causing rapid degradation of the CLH, etc. by inhibiting association of the CLH with the effectors which regulate the expression of BMPs, etc. Deactivators may be polypeptides, nucleic acids, carbohydrates, lipids, or derivatives thereof, or any other molecules which bind to and modulate the activity of said product.

"Treating a disease" - refers to administering a therapeutic substance effective to ameliorate symptoms associated with a disease, to lessen the severity or cure the disease, or to prevent the disease from occurring.

"***Detection***" – refers to a method of detection of a disease. This term may refer to detection of a predisposition to a disease.

"Probe" – the CLH nucleic acid sequence, or a sequence complementary therewith, when used to detect presence of other similar sequences in a sample. The detection is carried out by identification of hybridization complexes between the probe and the assayed sequence. The probe may be attached to a solid support or to a detectable label.

SUMMARY OF THE INVENTION

The present invention is based on the surprising finding that there exist in humans (and mice), several novel homologs of the chordin protein having a significant homology to the chordin protein, the homolog is a part of a longer sequence termed *"full sequence"*. The invention is further based on the surprising finding that there exist several splice variants to the full sequence which variants are naturally occurring sequences produced from the novel homolog through alternative splicing. Both the homolog and the variants of the full sequence are collectively termed as *"CLH"*.

The novel CLH (in SEQ ID NO: 1) is a homolog to the known chordins within the VWFC domain, named after the von-Willebrand factor (VWF) type C repeat, which is found 2-4 times in these multi-domain proteins. The VWF domain has a length of about 70 amino acids covering 10 well conserved cysteines. The presence of this region in complex-forming proteins leads to the assumption that the VWFC domain might be involved in forming larger protein complexes. The other variants to the full sequence (for which the homolog is a portion) (SEQ ID NO: 2-10) have 2, 3 or 4 VWF type repeats. SEQ ID NO. 2 and 6 also has a sequence coding for a signal sequence, while SEQ ID NO: 3 and 4 are predicted not to have such a signal sequence.

Thus the present invention provides by its first aspect, a novel isolated nucleic acid molecule comprising or consisting of the sequence of any one of SEQ ID NO: 1 to SEQ ID NO: 11, fragments of said sequence having at least 20 nucleic acids, or a molecule comprising a sequence having at least 70%, preferably 80%, and most preferably 90% or 95% identity to any one of SEQ ID NO:1 to SEQ ID NO: 11.

The present invention further provides a protein or polypeptide comprising or consisting of an amino acid sequence encoded by any of the above nucleic acid sequences, termed herein *"CLH product"*, for example, an amino acid sequence having the sequence as depicted in any one of SEQ ID NO: 12 to 22, fragments of the above amino acid sequence having a length of at least 10 amino acids, as well

as homologs of the amino acid sequences of any one of SEQ ID NO: 12 to 22 in which one or more of the amino acid residues has been substituted (by conservative or non-conservative substitution) added, deleted, or chemically modified.

The present invention further provides nucleic acid molecule comprising or consisting of a sequence which encodes the above amino acid sequences, (including the fragments and analogs of the amino acid sequences). Due to the degenerative nature of the genetic code, a plurality of alternative nucleic acid sequences, beyond SEQ ID NO:1 to SEQ ID NO: 11, can code for the amino acid sequence of the invention. Those alternative nucleic acid sequences which code for the same amino acid sequences codes by the sequences of SEQ ID NO: 1 to SEQ ID NO: 11 are also an aspect of the of the present invention.

The present invention further provides expression vectors and cloning vectors comprising any of the above nucleic acid sequences, as well as host cells transfected by said vectors.

The present invention still further provides pharmaceutical compositions comprising, as an active ingredient, said nucleic acid molecules, said expression vectors, or said protein or polypeptide.

By a second aspect, the present invention provides a nucleic acid molecule comprising or consisting of a non-coding sequence which is complementary to that of any one of SEQ ID NO: 1 to SEQ ID NO: 11, or complementary to a sequence having at least 70%, preferably 80%, most preferably 90% or 95% identity to said sequence or a fragment of said two sequences. The complementary sequence may be a DNA sequence which hybridizes with any one of the sequences of SEQ ID NO: 1 to SEQ ID NO: 11, or hybridizes to a portion of that sequence having a length sufficient to inhibit the transcription of the complementary sequence. The complementary sequence may be a DNA sequence which can be transcribed into an mRNA being an antisense to the mRNA transcribed from any one of SEQ ID NO: 1 to SEQ ID NO: 11 or into an mRNA which is an antisense to a fragment of the mRNA transcribed from any one of SEQ ID NO: 1 to SEQ ID NO: 11 which has a length sufficient to hybridize with the mRNA transcribed from any one of SEQ ID

NO: 1 to SEQ ID NO: 11, so as to inhibit its translation. The complementary sequence may also be the mRNA or the fragment of the mRNA itself. The pharmaceutical compositions of the invention (according to both aspects may be used for the treatment of a plurality of diseases.

In accordance with the present invention, it has been found that the CLH of the invention is located in astrocytes. As astrocytes are known to have a variety of physiological activities in maintaining normal brain physiology, such as in the secretion of active compounds, formation of the blood-brain barrier, metabolism of neurotransmitters and maintenance of the ionic balance of the extracellular space. Pharmaceutical compositions in accordance with the present invention may be used to treat diseases and pathological conditions which can be benefited by a modulation of astrocyte activity, such as the modulation of the cross-talk signals in the CNS during physiological and pathological conditions of the nervous system. Examples of such diseases are neuro-degenerative diseases caused by aging, infectious agents, by toxic substances or due to genetic causes. In addition, the pharmaceutical compositions may be used for the treatment of diseases, and pathological conditions involving abnormal development of the nervous system.

It has been postulated that chordin may be expressed by cells of the osteoblastic lineage to limit BMP actions in the osteoblast. This would be a critical function for a BMP binding protein since excessive BMP-4 has been implicated in pathogenesis of fibrodysplasia ossificans progressiva. Fibrodysplasia Ossificans Progressiva (FOP) is a rare genetic disease in which muscles, tendons, ligaments and other connective tissues may ossify into bone. BMPs can cause induction of noggin and chordin mRNA and protein levels in skeletan cells by trasncriptional mechanisms, and in turn these prevent the effect of BMPs in osteoblast in a negative-type feeback. The induction of these proteins by BMPs appears to be a mechanism to limit the BMP effect in bones. Existing therapies which are being investigated for their effectiveness in preventing heterotopic bone formation include BMP's inhibitors.

Considerable evidence exists supporting a role for TGF in morphogenesis,

in the regulation of endochondral ossification and in bone remodelling. TGF effect the proliferation and differentiation of osteoblastic cells *in vitro* and high levels of messenger RNA are expressed in the growth plate of fetal human long bones.

The CLH of the invention was found, by immunohystochemical methods to be localized in fetal-human bone .

Thus, the pharmaceutical compositions of the present invention may be used for the treatment of diseases and pathological conditions associated with osteoblastists or other diseases of mesanchimal origin. An example of such diseases is Fibrodysplasia Ossificans, as well as other diseases involving abnormal bone or cartilage formation, metabolism and/or destruction.

Furthermore, the CLH variances of the invention were mapped to chromosome 11q14 (genomic clone accession no. APOO 2010; AP001324; ACO118686).

The chromosomal location of the CLH gene is near several disorders of cartilage and bone formation, and thus, the pharmaceutical compositions of the invention may be used for the treatment or alleviation of the following diseases: Osteopetrosis, Autosomal Recessive (congenital disorder characterized also by development of abnormally dense bones).

High Bone Mass (HBM) - High bone mass can result from osteosclerosis (increased density of trabecular - spongy bone) and/or hyperostosis (thickening of cortical - compact bone from deposition of osseous tissue) along subperiosteal and/or endosteal surfaces), occurring focally or throughout the skeleton.

The pharmaceutical compositions of the invention may be used also for the treatment of osteoporosis pseudoglioma syndrome, autosomal recessive osteopetrosis, and isolated increased bone mass (high bone mass without other clinical features). The CLH of the invention may also be used for augmenting bone regeneration after injury, so as to speed up the healing process.

In accordance with the findings of the present invention, CLH of the invention is expressed in the placenta, and is localized in the uterus lining

(endometrium). It is known, that poor preparation of the endometrium (uterine lining) has been associated with abnormal pregnancies and a high rate of miscarriage, as well as other disorders of the female reproductive tract. Thus, the pharmaceutical compositions of the invention may be used for the support of a normal pregnancy, as well as for the treatment of abnormal pregnancies, recurring miscarriages, or the malfunction of the female reproductive tract.

Furthermore, the expression of CLH of the invention has also been found to be located in the mullerian epithel in the internal female ganglia (fallopian tubes, uterus, endocervix glands). The CLH of the invention can be used to regulate sexual differentiation, for example, by interaction with Mullerian inhibitory substances (MLS), a substance secreted by the testes, which causes the regression of the Mullerian duct system in females, leading to female sterility. In addition, the CLH of the invention may be used for the treatment of the Lawrence-Moon-Bardet-Biedl syndrome, a rare inherited condition, with variable expression, one of which is hypergenitalism (underdeveloped genitals).

In accordance with another finding of the invention, CLH was found to be expressed in tumors of the uterus, prostate and breast, indicating that CLH may be a proliferative agent on cell lines in general and tumor cell lines in particular. Thus, pharmaceutical compositions comprising an agent which decreased the expression or level of CLH, such as in anti-sense therapy, or antibodies, may be used for the treatment of these tumors.

The CLH of the invention is a hormone-responsive element, as it expressed in the mullerian epithelium, ductal epithelium of the breast, prostate, all of which are tissues under sexual hormonal control. Thus, since CLH is expressed in all estrogen target tissues (and some androgen target tissues), the pharmaceutical compositions of the invention may be used for hormonal regulation in such pathological conditions, involving non-normal amounts or a non-normal response to sexual hormones.

Pharmaceutical compositions of the invention may also be used for the treatment of cardiovascular disorders.

The nucleic acids of the invention may be used for therapeutic or diagnostic applications, for example. for the detection of the expression of CLH in various tissues, as mentioned above (for example, tumors, astrocytes, bone, tissues of the reproductive tract, etc.), and for the detection of any one of its diseases mentioned above. In addition, the ratio between the level of each of the chordin-like homologs to the other may also be indicative of a plurality of physiological or pathological conditions, for example, any one of the diseases mentioned above

The CLH gene of the invention was mapped to geonomic locus 11q14, a region containing many potential candidate bone diseases, neural system-related diseases, hormone-dependent diseases and developmental disorders. Thus, the detection of any of the CLH of the invention, as well as the detection of their amount or their ratio to each other, may be indicative to the presence of a disease, or a predisposition to a disease, or may be indicative of the severity of the disease. Furthermore, due to said association of the CLH of the invention with said disease, the pharmaceutical compositions of the invention (in connection with both aspects, i.e., both the nucleic acid sequence, the anti-sense, the amino acid sequence or the antibody) may be used for the treatment of said diseases or alleviation of some of their side effects.

The following is a list of diseases associated with the same geonomic locus as the CLH of the invention – which may be detected by the nucleic acid and amino acid sequences of the invention and the antibodies the invention and treated by the pharmaceutical compositions of the invention:

BONE RELATED DISEASES:

<u>Osteopetrosis, Autosomal Recessivea</u>

A rare hereditary disease characterized by extreme density and hardness and abnormal fragility of the bones with partial or complete obliteration of the marrow cavities. In this disorder there is a defective resorption of immature bone.

Osteoporosis-Pseudoglioma Syndrome; Oppg

A hereditary disease characterized by abnormally brittle, easily fractured bones, suggesting osteogenesis imperfecta.

High Bone Mass

High spinal bone mineral density

Osteoarthritis Susceptibility, Female-Specific

Somatotrophinoma, Acromegaly

A chronic disease of adults marked by enlargement of the bones of the extremities, face, and jaw that is caused by overactivity of the pituitary gland.

NERVOUS SYSTEM RELATED DISEASES

Pheochromocytoma, Familial Extraadrenal (Also Named Paragangliomas, Hereditary Extraadrenal)

A usually benign tumor of the adrenal medulla or the sympathetic nervous system in which the affected cells secrete increased amounts of epinephrine or norepinephrine. Disorder appears to have been due to a gene on 11q.

Tuberous Sclerosis 4

An inherited disorder of the skin and nervous system that is characterized typically by epilepsy and mental retardation, by a rash of the face resembling acne, and by multiple noncancerous tumors of the brain, kidney, retina, and heart failure, with radiographic evidence of cardiomegaly in all of them. Typical findings of tuberous sclerosis in the central nervous system, kidneys, heart, and liver.

Alexander Disease

This disorder, is characterized clinically by development of megalencephaly in infancy accompanied by progressive spasticity and dementia. In this disorder astrocytes show marked changes.

Hartnup Disorder

This disorder is characterized by a pellagra-like light-sensitive rash, cerebellar

ataxia, emotional instability, and aminoaciduria.

Spinal Muscular Atrophy With Respiratory Distress 1

Disorder that is characterized by the degeneration of motoneurons in the spinal cord resulting in muscular weakness and atrophy and that in some forms are fatal. neurogenic atrophy of skeletal muscle is observed .

Meckel Syndrome, Type 2; Mks2a

Syndrome inherited as an autosomal recessive trait and typically characterized by occipital encephalocele, microcephaly, cleft palate, polydactyly, and polycystic kidneys.

Schizophrenia Susceptibility Locus, Chromosome 11q-RELATED

Psychotic disorders usually characterized by withdrawal from reality, illogical patterns of thinking, delusions, and hallucinations, and accompanied in varying degrees by other emotional, behavioral, or intellectual disturbances. Schizophrenia, often associated with dopamine imbalances in the brain and defects of the frontal lobe, may have an underlying genetic cause.

DEVELOPMENTAL DISORDERS

Since Chordin play a role in patterning the early embryo development, Chordin-LM might involved in the following disorders:

Ebstein Anomaly

A congenital malformation of the heart that consists of downward placement of the tricuspid valve such that part of the right ventricle becomes incorporated into the pretricuspid chamber. Rearrangements of the long arm of chromosome 11 were described in patients with Ebstein anomaly.

Rutledge Lethal Multiple Congenital Anomaly Syndrome

External features, mesomelic dwarfism, micrognathia, V-shaped upper lip, microglossia, thick alveolar ridges, ambiguous genitalia, webbed neck, highly arched palate, clubfeet, fused fontanelles, inclusion cysts of the tongue, widely spaced nipples, and digital anomalies. Internal findings included oligopapillary renal hypoplasia, severe congenital heart defect, cerebellar hypoplasia, and

pulmonary, laryngeal, and gallbladder hypoplasia.

Bardet-Biedl Syndrome, Type 1; Bbs1

The Bardet-Biedl syndrome is characterized by mental retardation, pigmentary retinopathy, polydactyly, obesity, and hypogenitalism. The disorder is inherited as an autosomal recessive.

Targeted inactivation of chordin results in animals that display defects in inner and outer ear development. Therefore chordin-LM might be involved in hearing disorders such as the one linked to chromosome 11 - DEAFNESS, Autosomal Dominant Nonsyndromic Sensorineural 11; Dfna11.

The present invention also provides expression vectors comprising any one of the above defined complementary nucleic acid sequences and host cells transfected with said nucleic acid sequences or vectors, being complementary to those specified in the first aspect of the invention.

The invention also provides anti-CLH product antibodies, namely antibodies directed against the CLH product which specifically bind to said CLH product. Said antibodies are useful both for diagnostic and therapeutic purposes. For example said antibody may be as an active ingredient in a pharmaceutical composition as will be explained below.

The present invention also provides pharmaceutical compositions comprising, as an active ingredient, the nucleic acid molecules which comprise or consist of said complementary sequences, or of a vector comprising said complementary sequences. The pharmaceutical composition thus provides pharmaceutical compositions comprising, as an active ingredient, said anti-CLH product antibodies.

The pharmaceutical compositions comprising said anti-CLH product antibodies or the nucleic acid molecule comprising said complementary sequence, are suitable for the treatment of diseases and pathological conditions where a therapeutically beneficial effect may be achieved by neutralizing the CLH or decreasing the amount of the CLH product or blocking its binding to its target (for example its receptor), for example, by the neutralizing effect of the antibodies, or

by the decrease of the effect of the antisense mRNA in decreasing expression level of the CLH product. Examples of the diseases are any one of those mentioned above.

According to the third aspect of the invention the present invention provides methods for detecting the level of the transcript (mRNA) of said CLH product in a body fluid sample, or in a specific tissue sample or body fluid, for example, by use of probes comprising or consisting of said coding sequences (or complementary sequences); as well as methods for detecting levels of expression of said product in tissue, e.g. by the use of antibodies capable of specifically reacting with the above amino acid sequences. Detection of the level of the expression of the CLH of the invention may be indicative of a plurality of physiological or pathological conditions.

The method, according to this latter aspect, for detection of a nucleic acid sequence which encodes the CLH product in a biological sample, comprises the steps of:

(a) providing a probe comprising at least one of the nucleic acid sequence defined above;

(b) contacting the biological sample with said probe under conditions allowing hybridization of nucleic acid sequences thereby enabling formation of hybridization complexes;

(c) detecting hybridization complexes, wherein the presence of the complex indicates the presence of nucleic acid sequence encoding the CLH product in the biological sample.

The amount of hybridization complexes may be determined and calibrated by comparing it to a calibration scale in order to determine the amount of the nucleic acid sequence which enables the CLH product in the sample. The level of each of the sequences may be detected and either compared to the calibrated levels or to the level of each other, and said ratio may also be indicative to a plurality of pathological or physiological conditions.

By a preferred embodiment the probe is part of a nucleic acid chip used for detection purposes, i.e. the probe is a part of an array of probes each present in a known location on a solid support.

The nucleic acid sequence used in the above method may be a DNA sequence an RNA sequence, etc; it may be a coding or a sequence or a sequence complementary thereto (for respective detection of RNA transcripts or coding-DNA sequences). By quantization of the level of hybridization complexes and calibrating the quantified results it is possible also to detect the level of the transcript in the sample.

Methods for detecting mutations in the region coding for the CLH product are also provided, which may be methods carried-out in a binary fashion, namely merely detecting whether there is any mismatches between the normal CLH nucleic acid sequence and the one present in the sample, or carried-out by specifically detecting the nature and location of the mutation.

The present invention also concerns a method for detecting CLH product both for determining its presence, as well as its level or alterations in its level in a biological sample, comprising the steps of:

(a) contacting with said biological sample the antibody of the invention, thereby forming an antibody-antigen complex; and

(b) detecting said antibody-antigen complex

wherein the presence of said antibody-antigen complex correlates with the presence of CLH product in said biological sample.

The present invention also concerns a method for detecting anti-CLH antibodies in a biological sample comprising the steps of:

(a) contacting said biological sample with the product of the invention thereby forming an antibody-antigen complex; and

(b) detecting said antibody-antigen complex

wherein the presence of said antibody-antigen complex correlates with the presence of anti-CHL antibody in said biological sample.

In many cases, diseases are detected not by detecting the presence of the protein (product) which caused the disease, but rather by detecting the presence in a biological sample (such as blood or serum) of antibodies against such a product.

The method of detecting the presence of anti-CLH antibodies is intended to be used in such case.

The amount of the antibody-antigen complex can be quantitized, in order to determine the level of the CHL-product or the anti-CHL antibodies, as the case may be.

As explained above, the level of any one of the products may be compared to each other, and the ratio between the levels may be indicative to a plurality of physiological and pathological conditions. In addition, the indicative ratio may not be the ratio of the proteins themselves but rather the ratio of antibodies against the proteins.

By yet another aspect the invention also provides a method for identifying candidate compounds capable of modulating the activity of CLH product (being either activators or deactivators). The method includes:

(i) providing a protein or polypeptide comprising an amino acid sequence substantially as depicted in any one of SEQ ID NO: 12 to SEQ ID NO: 22, or a fragment of such a sequence;

(ii) contacting a candidate compound with said amino acid sequence;

(iii) comprising the physiological effect of the amino acid sequence in the presence and absence of said candidate compound and selecting those compounds which show a significant effect on said physiological activity.

The activity of the amino acid which should be changed by the modulator (being either the activator or deactivator) may be for example the binding of the CLH product to its receptor, the effect of CLH on BMPs expression or activity. Any modulator which changes such an activity has an infecting potential, as serving as an actuator or deactivator.

The present invention also concerns compounds identified by the above methods described above, which compound may either be an activator of the CLH product or a deactivator thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

In order to understand the invention and to see how it may be carried out in practice, a preferred embodiment will now be described, by way of non-limiting example only, with reference to the accompanying drawings, in which:

Fig. 1 is alignment of the CLH product of SEQ ID NO: 12 to known chordin protein, demonstrating the homology regions within these proteins. The alignment was performed using best-fit of GCG;

Fig. 2a is the alignment of the first splice variant (SEQ ID NO: 13) to the known chordin deposited in the Emb as gi 4808227;

Fig. 2b is the alignment of the first splice variant (SEQ ID NO: 13) to the known chordin deposited in the Emb under gi 3822218;

Fig. 2c is the alignment of the first splice variant (SEQ ID NO: 13) to the known chordin deposited in the Emb under gi 3800772;

Fig. 3a is the alignment of the second splice variant (SEQ ID NO: 14) with a known chordin deposited in the Emb under gi 4808227;

Fig. 3b is the alignment of the second splice variant (SEQ ID NO: 14) with a known chordin deposited in the Emb under gi 3822218;

Fig. 4a is the alignment of the third splice variant (SEQ ID NO: 15) with a known chordin deposited in the Emb under gi 4808227;

Fig. 4b is the alignment of the third splice variant (SEQ ID NO: 15) with a known chordin deposited in the Emb under gi 2731578;

Fig. 4c is the alignment of the third splice variant (SEQ ID NO: 16) with a known chordin deposited in the Emb under gi 2498235;

Fig. 4d is the alignment of the third splice variant (SEQ ID NO: 16) with a known chordin deposited in the Emb under gi 3822218;

Fig. 5 is multiple alignments of the sequences of the first four splice variants to several known chordins;.

Fig. 6 is the alignment of SEQ ID No. 16 to the known chordin deposited as gi 48082227;

Fig. 7 is the alignment of SEQ ID No. 16 to the known chordin deposited as gi 3822218;

Fig. 8 is the alignment of SEQ ID No. 16 to the known chordin deposited as gi 6753418;

Fig. 9 shows the alignment of SEQ ID No. 17 to the known chording deposited as gi 4808227;

Fig. 10 is the alignment of SEQ ID No. 17 to the known chordin deposited as gi 3822218;

Fig. 11 shows the alignment of SEQ ID No. 18 to the known chordin deposited as gi 4808222;.

Fig. 12 shows the alignment of SEQ ID No. 18 to the known chordin deposited as gi 3822218;

Fig. 13 shows the alignment of SEQ ID No. 19 to the known chordin deposited as gi 2731578;

Fig. 14 shows the alignment of SEQ ID No. 18 to the known chordin deposited as gi 3822218;

Fig. 15 shows the alignment of SEQ ID No. 20 to the known chordin deposited as gi 2731578;

Fig. 16 shows the alignment of SEQ ID No. 20 to the known chordin deposited as gi 382218;

Fig. 17 shows the alignment of SEQ ID No. 21 to the known chordin deposited as gi 2731578;

Fig. 18 shows the alignment of SEQ ID No. 21 to the the known chordin deposited as gi 3822218.

Fig. 19 shows multiple alignments of SEQ ID Nos. 12-21 (termed var 1-var 6, respectively) to each other;

Fig. 20 shows the alignment of SEQ ID No. 22 to the known chordin deposited as gi 480827.

Fig. 21 shows the alignment of SEQ ID No. 22 to the known chordin deposited as gi 6753418.

Fig. 22 shows a Northern blot analysis of CLH expression in: skeletal muscles, uterus, colon, small intestine, bladder, heart, stomach, prostate;

Fig. 23 shows a Western blot analysis of transfected COS-7 cells which express and secrete CHL;

Fig. 24 shows immunohistochemistry results with breast carcinoma (ductal and invasive ductal); prostate (carcinoma and benign prostate hyperplasia); bladder transitional epithelium; Mullerian Epithelium; uterus, bone Glioblastoma Multi-form (GBM); and

Fig. 25 shows Western blot analysis, or expression of CLH in human brain and bone tissues;

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Example I: CLH - nucleic acid sequence

The nucleic acid sequences of the invention include nucleic acid sequences which encode CLH product and fragments and analogs thereof. The nucleic acid sequences may alternatively be sequences complementary to the above coding sequence, or to a region of said coding sequence. The length of the complementary sequence is sufficient to avoid the expression of the coding sequence. The nucleic acid sequences may be in the form of RNA or in the form of DNA, and include messenger RNA, synthetic RNA and DNA, cDNA, and genomic DNA. The DNA may be double-stranded or single-stranded, and if single-stranded may be the coding strand or the non-coding (anti-sense, complementary) strand. The nucleic acid sequences may also both include dNTPs, rNTPs as well as non naturally occurring sequences. The sequence may also be a part of a hybrid between an amino acid sequence and a nucleic acid sequence.

In a general embodiment, the nucleic acid sequence has at least 70%, preferably 80% or 90% or 95% sequence identity with any one of the sequences identified as SEQ ID NO: 1 to SEQ ID NO: 11.

The nucleic acid sequences may include the coding sequence by itself. By another alternative the coding region may be in combination with additional coding sequences, such as those coding for fusion protein or signal peptides, in combination with non-coding sequences, such as introns and control elements, promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host, and/or in a vector or host environment in which the CLH nucleic acid sequence is introduced as a heterologous sequence.

The nucleic acid sequences of the present invention may also have the product coding sequence fused in-frame to a marker sequence which allows for purification of the CLH product. The marker sequence may be, for example, a

hexahistidine tag to provide for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., *et al. Cell* 37:767 (1984)).

Also included in the scope of the invention are fragments also referred to herein as oligonucleotides, typically having at least 20 bases, preferably 20-30 bases corresponding to a region of the coding-sequence nucleic acid sequence. The fragments may be used as probes, primers, and when complementary also as antisense agents, and the like, according to known methods.

As indicated above, the nucleic acid sequence may be substantially a depicted in any one of SEQ ID NO: 1 to SEQ ID NO: 11 or fragments thereof or sequences having at least 70%, preferably 70-80%, most preferably 90% or 95% identity to the above sequence. Alternatively, due to the degenerative nature of the genetic code, the sequence may be a sequence coding the amino acid sequence of any one of SEQ ID NO: 12 to SEQ ID NO: 22, or fragments or analogs of said amino acid sequence.

A. Preparation of nucleic acid sequences

The nucleic acid sequences may be obtained by screening cDNA libraries using oligonucleotide probes which can hybridize to or PCR-amplify nucleic acid sequences which encode the CLH products disclosed above. cDNA libraries prepared from a variety of tissues are commercially available and procedures for screening and isolating cDNA clones are well-known to those of skill in the art. Such techniques are described in, for example, Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Cold Spring Harbor Press, Plainview, N.Y. and Ausubel FM et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

The nucleic acid sequences may be extended to obtain upstream and downstream sequences such as promoters, regulatory elements, and 5' and 3'

untranslated regions (UTRs). Extension of the available transcript sequence may be performed by numerous methods known to those of skill in the art, such as PCR or primer extension (Sambrook *et al., supra*), or by the RACE method using, for example, the Marathon RACE kit (Clontech, Cat. # K1802-1).

Alternatively, the technique of *"restriction-site"* PCR (Gobinda *et al. PCR Methods Applic.* **2**:318-22, (1993)), which uses universal primers to retrieve flanking sequence adjacent a known locus, may be employed. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. *et al.*, *Nucleic Acids Res.* **16**:8186, (1988)). The primers may be designed using OLIGO(R) 4.06 Primer Analysis Software (1992; National Biosciences Inc, Plymouth, Minn.), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68-72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom, M. *et al., PCR Methods Applic.* **1**:111-19, (1991)) is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into a flanking part of the DNA molecule before PCR.

Another method which may be used to retrieve flanking sequences is that of Parker, J.D., *et al., Nucleic Acids Res.*, **19**:3055-60, (1991)). Additionally, one can use PCR, nested primers and PromoterFinder™ libraries to *"walk in"* genomic

DNA (PromoterFinder™; Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions. Preferred libraries for screening for full length cDNAs are ones that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred in that they will contain more sequences which contain the 5' and upstream regions of genes.

A randomly primed library may be particularly useful if an oligo d(T) library does not yield a full-length cDNA. Genomic libraries are useful for extension into the 5' nontranslated regulatory region.

The nucleic acid sequences and oligonucleotides of the invention can also be prepared by solid-phase methods, according to known synthetic methods. Typically, fragments of up to about 100 bases are individually synthesized, then joined to form continuous sequences up to several hundred bases.

B. Use of CLH nucleic acid sequence for the production of CLH products

In accordance with the present invention, nucleic acid sequences specified above may be used as recombinant DNA molecules that direct the expression of CLH products.

As will be understood by those of skill in the art, it may be advantageous to produce CLH product-encoding nucleotide sequences possessing codons other than those which appear in any one of SEQ ID NO: 1 to SEQ ID NO: 11 which are those which naturally occur in the human genome. Codons preferred by a particular prokaryotic or eukaryotic LHost (Murray, E. *et al. Nuc Acids Res.*, **17**:477-508, (1989)) can be selected, for example, to increase the rate of CLH product expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring sequence.

The nucleic acid sequences of the present invention can be engineered in order to alter a CLH product coding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or

expression of the product. For example, alterations may be introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to produce splice variants, etc.

The present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are also described in Sambrook, *et al., (supra)*.

The present invention also relates to host cells which are genetically engineered with vectors of the invention, and the production of the product of the invention by recombinant techniques. Host cells are genetically engineered (i.e., transduced, transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the expression of the CLH nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art.

The nucleic acid sequences of the present invention may be included in any one of a variety of expression vectors for expressing a product. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral

DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host. The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and related sub-cloning procedures are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate transcription control sequence (promoter) to direct mRNA synthesis. Examples of such promoters include: LTR or SV40 promoter, the *E.coli lac* or *trp* promoter, the phage lambda *PL* promoter, and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation, and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E.coli*.

The vector containing the appropriate DNA sequence as described above, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein. Examples of appropriate expression hosts include: bacterial cells, such as *E.coli*, *Streptomyces, Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila* and *Spodoptera* Sf9; animal cells such as CHO, COS, HEK 293 or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein. The invention is not limited by the host cells employed.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the CLH product. For example, when large

quantities of CLH product are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, multifunctional *E.coli* cloning and expression vectors such as *Bluescript*(R) (Stratagene), in which the CLH polypeptide coding sequence may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced; *pIN* vectors (Van Heeke & Schuster *J. Biol. Chem.* **264**:5503-5509, (1989)); *pET* vectors (Novagen, Madison WI); and the like.

In the yeast *Saccharomyces cerevisiae* a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used. For reviews, see Ausubel *et al. (supra)* and Grant *et al.*, (*Methods in Enzymology* **153**:516-544, (1987)).

In cases where plant expression vectors are used, the expression of a sequence encoding CLH product may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of *CaMV* (Brisson *et al., Nature* **310**:511-514. (1984)) may be used alone or in combination with the omega leader sequence from TMV (Takamatsu *et al., EMBO J.*, **6**:307-311, (1987)). Alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi *et al., EMBO J.* **3**:1671-1680, (1984); Broglie *et al., Science* **224**:838-843, (1984)); or heat shock promoters (Winter J and Sinibaldi R.M., *Results Probl. Cell Differ.*, **17**:85-105, (1991)) may be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. For reviews of such techniques, see Hobbs S. or Murry L.E. (1992) in McGraw Hill Yearbook of Science and Technology, McGraw Hill, New York, N.Y., pp 191-196; or Weissbach and Weissbach (1988) *Methods for Plant Molecular Biology*, Academic Press, New York, N.Y., pp 421-463.

CLH product may also be expressed in an insect system. In one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a

vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The CLH product coding sequence may be cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of CLH coding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect *S. frugiperda* cells or *Trichoplusia* larvae in which CLH protein is expressed (Smith *et al., J. Virol.* 46:584, (1983); Engelhard, E.K. *et al., Proc. Nat. Acad. Sci.* **91**:3224-7, (1994)).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, a CLH product coding sequence may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing CLH protein in infected host cells (Logan and Shenk, *Proc. Natl. Acad. Sci.* **81**:3655-59, (1984). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of a CLH protein coding sequence. These signals include the ATG initiation codon and adjacent sequences. In cases where CLH product coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf, D. *et al.*,

(1994) *Results Probl. Cell Differ.*, **20**:125-62, (1994); Bittner et al., *Methods in Enzymol* **153**:516-544, (1987)).

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., and Battey, I. (1986) Basic Methods in Molecular Biology). Cell-free translation systems can also be employed to produce polypeptides using RNAs derived from the DNA constructs of the present invention.

A host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the protein include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a *"pre-pro"* form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, 293, WI38, etc. have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express CLH product may be transformed using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the

introduced sequences. Resistant clumps of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M., *et al.*, *Cell* **11**:223-32, (1977)) and adenine phosphoribosyltransferase (Lowy I., *et al.*, *Cell* **22**:817-23, (1980)) genes which can be employed in *tk-* or *aprt-* cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, *dhfr* which confers resistance to methotrexate (Wigler M., *et al., Proc. Natl. Acad. Sci.* **77**:3567-70, (1980)); *npt*, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. *et al., J. Mol. Biol.*, **150**:1-14, (1981)) and *als* or *pat*, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine (Hartman S.C. and R.C. Mulligan, *Proc. Natl. Acad. Sci.* **85**:8047-51, (1988)). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate, GUS, and luciferase and its substrates, luciferin and ATP, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. *et. al., Methods Mol. Biol.*, 55:121-131, (1995)).

Host cells transformed with a nucleotide sequence encoding CLH product may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The product produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing nucleic acid sequences encoding CLH product can be designed with signal sequences which direct secretion of CLH product through a prokaryotic or eukaryotic cell membrane.

CLH product may also be expressed as a recombinant protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle, Wash.). The inclusion of a protease-cleavable polypeptide linker sequence between the purification domain and CLH protein is useful to facilitate purification. One such expression vector provides for expression of a fusion protein compromising a CLH polypeptide fused to a polyhistidine region separated by an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography, as described in Porath, *et al.*, *Protein Expression and Purification,* **3**:263-281, (1992)) while the enterokinase cleavage site provides a means for isolating CLH polypeptide from the fusion protein. *pGEX* vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to ligand-agarose beads (e.g., glutathione-agarose in the case of GST-fusions) followed by elution in the presence of free ligand.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, or other methods, which are well know to those skilled in the art.

The CLH products can be recovered and purified from recombinant cell cultures by any of a number of methods well known in the art, including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

C. Diagnostic applications utilizing nucleic acid sequences

The nucleic acid sequences of the present invention may be used for a variety of diagnostic purposes. The nucleic acid sequences may be used to detect and quantitate expression of CLH in patient's cells, e.g. biopsied tissues, by detecting the presence of mRNA coding for CLH product. Alternatively, the assay may be used to detect soluble CLH in the serum or blood. This assay typically involves obtaining total mRNA from the tissue or serum and contacting the mRNA with a nucleic acid probe. The probe is a nucleic acid molecule of at least 20 nucleotides, preferably 20-30 nucleotides, capable of specifically hybridizing with a sequence included within the sequence of a nucleic acid molecule encoding CLH under hybridizing conditions, detecting the presence of mRNA hybridized to the probe, and thereby detecting the expression of CLH. This assay can be used to distinguish between absence, presence, and excess expression of CLH product and to monitor levels of CLH expression during therapeutic intervention.

The invention also contemplates the use of the nucleic acid sequences as a diagnostic for diseases resulting from inherited defective CLH sequences. These sequences can be detected by comparing the sequences of the defective (i.e., mutant) CLH coding region with that of a normal coding region. Association of the sequence coding for mutant CLH product with abnormal CLH product

activity may be verified. In addition, sequences encoding mutant CLH products can be inserted into a suitable vector for expression in a functional assay system (e.g., colorimetric assay, complementation experiments in a CLH protein deficient strain of HEK293 cells) as yet another means to verify or identify mutations. Once mutant genes have been identified, one can then screen populations of interest for carriers of the mutant gene.

Individuals carrying mutations in the nucleic acid sequence of the present invention may be detected at the DNA level by a variety of techniques. Nucleic acids used for diagnosis may be obtained from a patient's cells, including but not limited to such as from blood, urine, saliva, placenta, tissue biopsy and autopsy material. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki, *et al.*, *Nature* **324**:163-166, (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid of the present invention can be used to identify and analyze mutations in the gene of the present invention. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype.

Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA of the invention or alternatively, radiolabeled antisense DNA sequences of the invention. Sequence changes at specific locations may also be revealed by nuclease protection assays, such RNase and S1 protection or the chemical cleavage method (e.g. Cotton, *et alProc. Natl. Acad. Sci.* USA, **85**:4397-4401, (1985)), or by differences in melting temperatures. *"Molecular beacons"* (Kostrikis L.G. *et al.*, Science **279**:1228-1229, (1998)), hairpin-shaped, single-stranded synthetic oligo- nucleotides containing probe sequences which are complementary to the nucleic acid of the present invention, may also be used to detect point mutations or other sequence changes as well as monitor expression levels of CLH product. Such diagnostics would be particularly useful for prenatal testing.

Another method for detecting mutations uses two DNA probes which are designed to hybridize to adjacent regions of a target, with abutting bases, where the region of known or suspected mutation(s) is at or near the abutting bases. The two probes may be joined at the abutting bases, e.g., in the presence of a ligase enzyme, but only if both probes are correctly base paired in the region of probe junction. The presence or absence of mutations is then detectable by the presence or absence of ligated probe.

Also suitable for detecting mutations in the CLH product coding sequence are oligonucleotide array methods based on sequencing by hybridization (SBH), as described, for example, in U.S. Patent No. 5,547,839. In a typical method, the DNA target analyte is hybridized with an array of oligonucleotides formed on a microchip. The sequence of the target can then be *"read"* from the pattern of target binding to the array.

D. Gene mapping utilizing nucleic acid sequences

The nucleic acid sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 20-30 bp) from the CLH cDNA. Computer analysis of the 3' untranslated region is used to rapidly select primers that do not span more than one exon in the genomic DNA, which would complicate the amplification process. These primers are then used for PCR screening of somatic cell hybrids

containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids or using instead radiation hybrids are rapid procedures for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 50 or 60 bases. For a review of this technique, see Verma *et al., Human Chromosomes: a Manual of Basic Techniques*, (1988) Pergamon Press, New York.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in the OMIM database (Center for Medical Genetics, Johns Hopkins University, Baltimore, MD and National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD). The OMIM gene map presents the cytogenetic map location of disease genes and other expressed genes. The OMIM database provides information on diseases associated with the chromosomal location. Such associations include the results of linkage analysis mapped to this interval, and the correlation of translocations and other chromosomal aberrations in this area with the advent of various diseases, for example, those mentioned in connection with the pharmaceutical compositions of the invention.

E. Therapeutic applications of nucleic acid sequences

Nucleic acid sequences of the invention may also be used for therapeutic purposes. Turning first to the second aspect of the invention (i.e. inhibition of expression of CLH), expression of CLH product may be modulated through antisense technology, which controls gene expression through hybridization of complementary nucleic acid sequences, i.e. antisense DNA or RNA, to the control, 5' or regulatory regions of the gene encoding CLH product. For example, the 5' coding portion of the nucleic acid sequence sequence which codes for the product of the present invention is used to design an antisense oligonucleotide of from about 10 to 40 base pairs in length. Oligonucleotides derived from the transcription CLHt site, e.g. between positions -10 and +10 from the CLHt site, are preferred. An antisense DNA oligonucleotide is designed to be complementary to a region of the nucleic acid sequence involved in transcription (Lee *et al., Nucl. Acids, Res.*, **6**:3073, (1979); Cooney et al., Science **241**:456, (1988); and Dervan *et al., Science* **251**:1360, (1991)), thereby preventing transcription and the production of the CLH products. An antisense RNA oligonucleotide hybridizes to the mRNA in vivo and blocks translation of the mRNA molecule into the CLH products (Okano *J. Neurochem.* **56**:560, (1991)). The antisense constructs can be delivered to cells by procedures known in the art such that the antisense RNA or DNA may be expressed *in vivo*. The antisense may be antisense mRNA or DNA sequence capable of coding such antisense mRNA. The antisense mRNA or the DNA coding thereof can be complementary to the full sequence of nucleic acid sequences coding to the CLH protein or to a fragment of such a sequence which is sufficient to inhibit production of a protein product.

Turning now to the first aspect of the invention, i.e. expression of CLH, expression of CLH product may be increased by providing coding sequences for coding for said product under the control of suitable control elements ending its expression in the desired host.

The nucleic acid sequences of the invention may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

The products of the invention as well as any activators and deactivators compounds (see below) which are polypeptides, may also be employed in accordance with the present invention by expression of such polypeptides *in vivo*, which is often referred to as *"gene therapy."* Cells from a patient may be engineered with a nucleic acid sequence (DNA or RNA) encoding a polypeptide *ex vivo*, with the engineered cells then being provided to a patient to be treated with the polypeptide. Such methods are well-known in the art. For example, cells may be engineered by procedures known in the art by use of a retroviral particle containing RNA encoding a polypeptide of the present invention.

Similarly, cells may be engineered in vivo for expression of a polypeptide *in vivo* by procedures known in the art. As known in the art, a producer cell for producing a retroviral particle containing RNA encoding the polypeptide of the present invention may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo*. These and other methods for administering a product of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention. For example, the expression vehicle for engineering cells may be other than a retrovirus, for example, an adenovirus which may be used to engineer cells *in vivo* after combination with a suitable delivery vehicle.

Retroviruses from which the retroviral plasmid vectors mentioned above may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus.

The retroviral plasmid vector is employed to transduce packaging cell lines to form producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the *PE501*, *PA317*, *psi-2*, *psi-AM*, *PA12*, *T19-14X*, *VT-19-17-H2*, *psi-CRE*, *psi-CRIP*, *GP+E-86*, *GP+envAm12*, and *DAN* cell lines as described in Miller (*Human Gene Therapy*, Vol. **1**, pg. 5-14, (1990)). The vector may transduce the packaging cells through any means known in the art. Such means include, but are not limited to, electroporation, the use of liposomes, and $CaPO_4$ precipitation. In one alternative, the retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host.

The producer cell line generates infectious retroviral vector particles which include the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed, to transduce eukaryotic cells, either in vitro or in vivo. The transduced eukaryotic cells will express the nucleic acid sequence(s) encoding the polypeptide. Eukaryotic cells which may be transduced include, but are not limited to, embryonic stem cells, embryonic carcinoma cells, as well as hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

The genes introduced into cells may be placed under the control of inducible promoters, such as the radiation-inducible Egr-1 promoter, (Maceri, H.J., *et al., Cancer Res.*, **56**(19):4311 (1996)), to stimulate CLH production or antisense inhibition in response to radiation, eg., radiation therapy for treating tumors.

Example II. CLH product

The substantially purified CLH product of the invention has been defined above as the product coded from the nucleic acid sequence of the invention. Preferably the amino acid sequence is an amino acid sequence having at least 70%, preferably at least 80% or 90% or 95% identity to the sequence identified as any one of SEQ ID NO: 12 to SEQ ID NO: 22. The protein or polypeptide may

be in mature and/or modified form, also as defined above. Also contemplated are protein fragments having at least 10 contiguous amino acid residues, preferably at least 10-20 residues, derived from the CLH product.

The sequence variations are preferably those that are considered conserved substitutions, as defined above. Thus, for example, a protein with a sequence having at least 80%, preferably 90% sequence identity with the protein identified as any one of SEQ ID NO: 12 to SEQ ID NO: 22, preferably by utilizing conserved substitutions as defined above is also part of the invention. In a more specific embodiment, the protein has or contains the sequence identified as any one of SEQ ID NO: 12 to SEQ ID NO: 22. The CLH product may be (i) one in which one or more of the amino acid residues in a sequence listed above are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue), or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the CLH product is fused with another compound, such as a compound to increase the half-life of the protein (for example, polyethylene glycol (PEG)), or a moiety which serves as targeting means to direct the protein to its target tissue or target cell population (such as an antibody), or (iv) one in which additional amino acids are fused to the CLH product. Such fragments, variants and derivatives are deemed to be within the scope of those skilled in the art from the teachings herein.

A. Preparation of CLH product

Recombinant methods for producing and isolating the CLH product, and fragments of the protein are described above.

In addition to recombinant production, fragments and portions of CLH product may be produced by direct peptide synthesis using solid-phase techniques (cf. Stewart *et al.*, (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco; Merrifield J., *J. Am. Chem. Soc.*, **85**:2149-2154, (1963)). In vitro peptide synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems

431A Peptide Synthesizer (Perkin Elmer, Foster City, Calif.) in accordance with the instructions provided by the manufacturer. Fragments of CLH product may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

B. Therapeutic uses and compositions utilizing the CLH product

The CLH product of the invention is generally useful in treating diseases and disorders which are characterized by a lower than normal level of CLH expression, and or diseases which can be cured or ameliorated by raising the level of the CLH product, even if the level is normal.

Typically these diseases are in CLH products or fragments and may be administered by any of a number of routes and methods designed to provide a consistent and predictable concentration of compound at the target organ or tissue. The product-containing compositions may be administered alone or in combination with other agents, such as stabilizing compounds, and/or in combination with other pharmaceutical agents such as drugs or hormones.

CLH product-containing compositions may be administered by a number of routes including, but not limited to oral, intravenous, intramuscular, transdermal, subcutaneous, topical, sublingual, or rectal means as well as by nasal application. CLH product-containing compositions may also be administered via liposomes. Such administration routes and appropriate formulations are generally known to those of skill in the art.

The product can be given via intravenous or intraperitoneal injection. Similarly, the product may be injected to other localized regions of the body. The product may also be administered via nasal insufflation. Enteral administration is also possible. For such administration, the product should be formulated into an appropriate capsule or elixir for oral administration, or into a suppository for rectal administration.

The foregoing exemplary administration modes will likely require that the product be formulated into an appropriate carrier, including ointments, gels,

suppositories. Appropriate formulations are well known to persons skilled in the art.

Dosage of the product will vary, depending upon the potency and therapeutic index of the particular polypeptide selected.

A therapeutic composition for use in the treatment method can include the product in a sterile injectable solution, the polypeptide in an oral delivery vehicle, the product in an aerosol suitable for nasal administration, or the product in a nebulized form, all prepared according to well known methods. Such compositions comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The product of the invention may also be used to modulate endothelial differentiation and proliferation as well as to modulate apoptosis either *ex vivo* or *in vitro*, for example, in cell cultures.

Example III. Screening methods for activators and deactivators (inhibitors)

The present invention also includes an assay for identifying molecules, such as synthetic drugs, antibodies, peptides, or other molecules, which have a modulating effect on the activity of the CLH product, e.g. activators or deactivators of the CLH product of the present invention. Such an assay comprises the steps of providing an CLH product encoded by the nucleic acid sequences of the present invention and determining its physiological activity on the target in the presence and absence of one or more candidate molecules to determine the candidate molecules. Those molecules which are modulating effect on the activity of the CLH product are selected as likely candidates for activators and deactivators.

CLH product, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any of a variety of drug screening techniques. The fragment employed in such a test may be free in solution, affixed to a solid support, borne on a cell membrane or located

intracellularly. The formation of binding complexes, between CLH product and the agent being tested, may be measured. Alternatively, the activator or deactivator may work by serving as agonist or antagonist, respectively, of the CLH receptor and their effect may be determined in connection with the receptor.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the CLH product is described in detail by Geysen in PCT Application WO 84/03564, published on Sep. 13, 1984. In summary, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with the full CLH product or with fragments of CLH product and washed. Bound CLH product is then detected by methods well known in the art. Substantially purified CLH product can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

Antibodies to the CLH product, as described in Example IV below, may also be used in screening assays according to methods well known in the art. For example, a *"sandwich"* assay may be performed, in which an anti-CLH antibody is affixed to a solid surface such as a microtiter plate and CLH product is added. Such an assay can be used to capture compounds which bind to the CLH product. Alternatively, such an assay may be used to measure the ability of compounds to influence with the binding of CLH product to the CLH receptor [I and then select those compounds which effect the binding.

Example IV. Anti-CLH antibodies

A. Synthesis

In still another aspect of the invention, the purified CLH product is used to produce anti-CLH antibodies which have diagnostic and therapeutic uses related to the activity, distribution, and expression of the CLH product, in particular

therapeutic applications mentioned in connection with the pharmaceutical composition aspect of the invention.

Antibodies to CLH product may be generated by methods well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, humanized, single chain, Fab fragments and fragments produced by an Fab expression library. Antibodies, i.e., those which inhibit dimer formation, are especially preferred for therapeutic use.

A fragment CLH product for antibody induction does not require biological activity but have to feature immunological activity; however, the protein fragment or oligopeptide must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids of the sequences specified in SEQ ID NO: 12 to SEQ ID No. 22. Preferably they should mimic a portion of the amino acid sequence of the natural protein and may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CLH protein amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule. Procedures well known in the art can be used for the production of antibodies to CLH product.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, etc may be immunized by injection with CLH product or any portion, fragment or oligopeptide which retains immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include but are not limited to Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are potentially useful human adjuvants.

Monoclonal antibodies to CLH protein may be prepared using any technique which provides for the production of antibody molecules by continuous

cell lines in culture. These include but are not limited to the hybridoma technique originally described by Koehler and Milstein (*Nature* **256**:495-497, (1975)), the human B-cell hybridoma technique (Kosbor *et al., Immunol. Today* **4**:72, (1983); Cote *et al., Proc. Natl. Acad. Sci.* **80**:2026-2030, (1983)) and the EBV-hybridoma technique (Cole, *et al., Mol. Cell Biol.* 62:109-120, (1984)).

Techniques developed for the production of *"chimeric antibodies"*, the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can also be used (Morrison *et al., Proc. Natl. Acad. Sci.* **81**:6851-6855, (1984); Neuberger *et al., Nature* **312**:604-608, (1984); Takeda *et al., Nature* **314**:452-454, (1985)). Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778) can be adapted to produce single-chain antibodies specific for the CLH protein.

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi *et al.* (*Proc. Natl. Acad. Sci.* **86**:3833-3837, 1989)), and Winter G and Milstein C., (*Nature* **349**:293-299, (1991)).

Antibody fragments which contain specific binding sites for CLH protein may also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse W.D. *et al., Science* 256:1275-1281, (1989)).

B. Diagnostic applications of antibodies

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established

specificities are well known in the art. Such immunoassays typically involve the formation of complexes between CLH product and its specific antibody and the measurement of complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two noninterfering epitopes on a specific CLH product is preferred, but a competitive binding assay may also be employed. These assays are described in Maddox D.E., *et al.*, (*J. Exp. Med.* **158**:1211, (1983)).

Antibodies which specifically bind CLH product are useful for the diagnosis of conditions or diseases characterized by over or under expression of CLH. Alternatively, such antibodies may be used in assays to monitor patients being treated with CLH product, its activators, or its deactivators. Diagnostic assays for CLH protein include methods utilizing the antibody and a label to detect CLH product in human body fluids or extracts of cells or tissues. The products and antibodies of the present invention may be used with or without modification. Frequently, the proteins and antibodies will be labeled by joining them, either covalently or noncovalently, with a reporter molecule. A wide variety of reporter molecules are known in the art.

A variety of protocols for measuring CLH product, using either polyclonal or monoclonal antibodies specific for the respective protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescent activated cell sorting (FACS). As noted above, a two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CLH product is preferred, but a competitive binding assay may be employed. These assays are described, among other places, in Maddox, *et al. (supra)*. Such protocols provide a basis for diagnosing altered or abnormal levels of CLH product expression. Normal or standard values for CLH product expression are established by combining body or cell extracts taken from normal subjects, preferably human, with antibody to CLH product under conditions suitable for complex formation which are well known in the art. The amount of standard complex formation may be quantified

by various methods, preferably by photometric methods. Then, standard values obtained from normal samples may be compared with values obtained from samples from subjects potentially affected by disease. Deviation between standard and subject values establishes the presence of disease state.

The antibody assays are useful to determine the level of CLH present in a body fluid sample, in order to determine whether it is being overexpressed or underexpressed in the tissue, or as an indication of how CLH levels are responding to drug treatment.

Another alternative is to determine the presence and/or level of naturally occurring anti-CLH antibodies in a sample, such as blood or serum. Many times diseases are identified by detecting the presence or level of antibodies against a specific product. For the detection of such naturally occurring anti-CLH antibodies, the sample may be contacted with the product of the invention, for example as depicted in any one of SEQ ID NO: 5 to SEQ ID NO: 8, or with an antigenic fragment thereof, and the presence or level of antibody-antigen complexes may be determined by methods well known in the art.

C. Therapeutic uses of antibodies

In addition to their diagnostic use the antibodies may have a therapeutical utility in blocking or decreasing the activity of the CLH product in pathological conditions where beneficial effect can be achieved by such a decrease.

The antibody employed is preferably a humanized monoclonal antibody, or a human Mab produced by known globulin-gene library methods. The antibody is administered typically as a sterile solution by IV injection, although other parenteral routes may be suitable. Typically, the antibody is administered in an amount between about 1-15 mg/kg body weight of the subject. Treatment is continued, e.g., with dosing every 1-7 days, until a therapeutic improvement is seen.

Although the invention has been described with reference to specific methods and embodiments, it is appreciated that various modifications and changes may be made without departing from the invention.

Example V: Experimental Procedures

A. <u>RNA Purification and cDNA Synthesis</u>

Total RNA was extracted from different human tissues using Tri-Reagent System (Molecular Research Center, Inc., Cincinnati, OH).

Poly (A) RNA was isolated from total RNA using Dynabeads mRNA Direct Kit (Dynal, Oslo, Norway).

Synthesis of first-strand cDNA was carried out using Oligo(dT)^{15} (Promega, Medison, WI), Superscript II or ThermoScript RNase H$^-$ Reverse transcriptase (Gibco/BRL, Gaithersburg, MD), Rnasin (Promega, Medison, WI) and dNTP's (Gibco/BRL, Gaithersburg, MD).

B. <u>RACE analysis of 5' and 3' ends of LM</u>

5' and 3' RACE (rapid amplification of cDNA ends) analysis was performed on poly A RNA from human placenta tissue using the Marathon cDNA Amplification Kit (Clontech). Adaptor-ligated double-stranded cDNA libraries were prepared essentially as suggested by the manufacturer. Superscript II Reverse Transcriptase (Gibco/BRL, Gaithersburg, MD) was used for the first strand synthesis. First round PCR was performed on these libraries for 30 cycles, using the Expand Long Template PCR System (Boerhringer-Mannheim, Germany). A nested PCR approach was used to isolate 5' and 3' RACE products.

C. <u>Polymerase Chain Reaction (PCR)</u>:

PCR was performed using either Taq DNA polymerase or Expand Long Template PCR system (Roche) pretreated with Taq Start Antibody (Hot

start). As a template we used cDNA from different tissues. The PCR reaction on PTC-225 (MJ Research, Inc.). PCR products were analyzed on an automated DNA sequencer ABI Prizem 310 Genetic Analyzer (Perkin Elmer).

D. Northern blot

Random primer DNA labeling was performed using ^{32}P and MIT (Biological Industries Co., Beit Haemek LTD). Chordin-LM probes used were a product of the sense primer:

5'- GAAAGCCTGTGTGCATGGCGG-3'

and the anti-sense primer:

5'-AGCTCATATCTGCAACTGTTAGG-3'.

The membranes used were from Human Muscle Multiple Tissue Northern Blot (MTN™, Clontech).

E. Expressiont of GST-LM

The PCR product was cloned into plasmid PGEX-6p (Pharmacia Biotech) and expressed in E.coli DH5-alfa as a fusion protein with GST. Expression, purification and detection of the fusion protein GST- LM was performed following the manufacturer's instructions.

F. Preparation of Antibodies

The anti- LM was prepared by immunizing rabbits with the purified fusion GST-LM.

G. Immunohistochemistry

Immunohistochemical staining was performed using Histostain plus Kit (Zymed Laboratories Inc.). Different human micron sections were prepared using a R. Gung microtome and fixed on superfrost plus slides with 2% Tespa. Deparaffinization was performed in xylene for 10 min.

Dehydration was performed three times in absolute ethanol and once 95% ethanol. The slides were washed in DDW and then icubated with 3% H_2O_2 for 5 min. Subsequently, the slide were washed in DDW and twice in 0.05M TrisHCl pH 7.6 (Optimax wash Buffer, BioGenex). The rest of the procedure was performed following the manufacturer's instructions.

H. Expression Plasmids:

The variants were cloned into pCDNA3 mammalian expression vector (Invitrogen).

I. Transfection experiments:

Chordin-LM was transiently expressed in COS-7 cell line (ATCC). The transfection of the expression vector into COS-7 was done by the FuGENE™ 6 method according to the manufacturer's instructions (Boeringer Mannheim).

J. Western blot

Protein samples were separated by SDS-PAGE, electrophoretically transferred to a nitrocellulose membrane (Pharmacia Biotech), and subjected to immunodetection using the immunized sera as a primary antibody and peroxidase-conjugated Gout Anti Rabbit IgG (Jackson Immunoresearch Laboratories, Inc.). Proteins were visualized with enhanced chemiluminescence system (Pierce).

EXAMPLE VI: Expression pattern of CLH

To begin characterization of CLH RNA expression, Northern blot analysis was performed. The membranes used were from Human Muscle Multiple Tissue Northern Blot (MTN™, Clontech). CLH mRNA of 2.3kb was detected at

significantly high levels in uterus, and also in colon, bladder, heart, stomach and prostate as shown in Fig. 22.

Expression of CLH mRNA was also found in different human cDNA tissues, such as: testis, placenta, brain, bone marrow, ovary, fetal lung, fetal brain. (data not shown).

EXAMPLE VII: Generation of CLH specific antibodies:

In order to generate antibodies against CLH, DNA fragment containing Chordin-like variant 1 was PCR-amplified and cloned into pGEX-6p vector (Pharmacia Biotech). Using the glutathione S- transferase (GST) gene fusion system, CLH fused to GST was expressed, purified and detected on SDS – PAGE. Large scale of CLH fused protein was prepared to immunized rabbits. Sera before the rabbits immunization was collected (reffered as the pre-immuned Ab's) and also following serial rabbit immunization with the purified fusion GST-LM (reffered as the anti-LM Ab's). The antibodies thus produced were used for immunohistochemical studies.

EXAMPLE VIII: Expression and secretion of CLH in mammalian cell line

As mentioned before, CLH SEQ ID No. 16 contains predicted signal peptide. In order to validate secretion of the protein, DNA fragment containing CLH SEQ ID No. 5 was PCR-amplified and cloned into pCDNA3 mammalian expression vector (Invitrogen).

COS-7 cells were transfected with pCDNA3 carrying CLH gene or with pCDNA3 alone. After incubating for 48 hr and 72 hr, mediums from the transfected cells were collected and protein extraction from the cells was performed. Protein samples from both the cell lysate and the medium were separated by SDS-PAGE, electrophoretically transferred to a nitrocellulose membrane and subjected to immunodetection using the anti- LM Ab's of example VII . The expression and the secretion of the CLH SEQ ID No. 13

variant 1 molecule is presented in figure 23.

As shown in Fig 22 , lanes A3, B3, C3, COS7 untransfected cells (referred to as *Mock)* do not express CLH endogenously. CLH was over expressed only in the cells transfected with pCDNA3 carrying CLH gene Fig 23 , lane C1 and not in the cells transfected with pCDNA3 Fig 23 , lane C2. Moreover, high levels of secreted protein were detected in the medium of CLH transfected cells following 48hr and 72 hr (Fig22 Lanes A1 and lane B1 respectively), and not in the cells transfected with pCDNA3 (Fig22 Lanes A2 and lane B2 respectively).

EXAMPLE IX: Immunohistochemical localization of CLH protein in different human tissues:

Immunohistochemical staining was performed on different human micron sections using the anti-LM antibodies (Fig 24 right column letters with prime) indicated compared to the pre-immune rabbit's serum (Fig 24, left columns, indicated in normal letters). CLH was found to be expressed in different epithelial tissues (Fig.24 a', b', c', d', e', f', g') and localized mainly in the secreting cells.

Expression of CLH was detected in ductal epithelium of the breast. Breast carcinoma was positively stained both in the regions of ductal carcinoma (Fig. 24 a') in situ (DCIS) and of invasive ductal carcinoma (Fig. 24b). Secreting cells in benign prostatic hyperplasia (BPH) and prostate carcinoma sections were also positively stained Fig. 24, c', d', respectively.

CLH was localized to the transitional epithelium in the bladder (Fig 24 e'). The internal female genitalia (fallopian tube, endocervical glands and the uterus) which evolved from the same embryonic precursor - the mullerian duct, showed positive staining (Fig.24, e'). Expression of CLH was localized in the lining epithelium of the fallopian tube (Fig 24, f'), in the endocervical glands (Fig 24, g') and in the normal and endometrial carcinoma of the uterus (Fig. 24, h' and i', respectively). However, in the region of the mucinous metaplasia in the endometrial carcinoma, negative staining of CLH was observed

(Fig 24, j').

CLH was localized not only in epithelial tissues as mentioned above, but also in osteoblasts in the fetal bone of thigh (Fig 24k').

Positive staining of CLH was also detected in activated astrocyte (referred as Gemistocyte , Fig 24l') in Glioblastoma Multiforme-GBM (brain tumor) but not in oligodendroglia (Fig 24l negative staining).

EXAMPLE X: CLH protein expression - Western Blot Analysis

CLH was detected in different tissues by Western blot analysis using anti-LM Ab's. As shown in Fig. 25, CLH is expressed in the brain and bone tumor (Fig 25A and 2B respectively). CLH in the transfection medium (experiment described in details previously), served as a positive control (Fig 25 refered as positive control). Multiple bands in the Western blot analysis may reflect either alternative splicing products of a single gene or post translational modifications (PTM) of CLH.